DOCUMENT RESUME

ED 425 172 TM 029 214

AUTHOR Tananis, Cynthia A.

TITLE Beyond Data: Evaluation as Discursive Practice within Math/Science Education Reform.

PUB DATE 1998-07-00

NOTE 55p.; Paper presented at the Annual Consortium for Research on Educational Accountability and Teacher Evaluation National Evaluation Institute (7th, Denver, CO, July 16-18, 1998).

PUB TYPE Information Analyses (070) -- Reports - Descriptive (141) -- Speeches/Meeting Papers (150)

EDRS PRICE MF01/PC03 Plus Postage.

DESCRIPTORS *Cooperation; *Educational Change; Elementary Secondary Education; *Evaluation Methods; *Mathematics Education; Organizational Development; Participative Decision Making; Program Evaluation; *Science Education

IDENTIFIERS *Discourse; External Evaluation; Stakeholders

ABSTRACT

The Regional Math/Science Collaborative (RMSC) of Southwestern Pennsylvania is a grassroots approach to strengthening math and science education through regional planning and evaluation. The Collaborative consists of stakeholders representing educators, students, parents, university faculty, and business and community members. External evaluation with the assistance of the University of Pittsburgh was planned to assist continuing regional discourse regarding mathematics and science education by developing a database to provide meaning and context for the indicators identified for the project. Because the issues inherent in participative evaluation are complex and challenging, a literature review was conducted to explore evaluation approaches. Three main strands of participative evaluation and many other variations of stakeholder evaluation were identified. These approaches differed in some key areas, but shared common elements, such as stakeholder involvement and the importance of dialogue between stakeholders and the external evaluator. A form of dialogue involving collective reflection and action characterized as discursive practice was planned for the evaluation. Evaluation based on discursive practice recognizes that organizations learn through joint discussion and interpretation of events and through gradual changes in the assumptions, symbols, and values of participants. It is the integral use of stakeholder dialogue, through iterative deliberation, that characterizes discursive practice. Through this form of participation, evaluation can support organizational development and fuller evaluation utilization, while building an institutionalized evaluation capacity. (Contains 94 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 425 172

Beyond Data: Evaluation as Discursive Practice Within Math/Science Education Reform

Cynthia A. Tananis

5p26 FQ – School of Education
University of Pittsburgh
Pittsburgh, PA 15260
Voice: (412) 648-7171
Fax: (412) 648-1784
Email: tananis+@pitt.edu

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Cynthia Tananis

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

TM029214

Paper presented at the Seventh Annual National Evaluation Institute, Denver, CO, July 16-18, 1998. Consortium for Research on Educational Accountability and Teacher Evaluation (CREATE)

Regional Math/Science Collaborative: A Regional Approach to Strengthening Math and Science Education

The Regional Math/Science Collaborative of Southwestern Pennsylvania (RMSC) is a grassroots approach to strengthening math and science education through regional planning and action. The Collaborative consists of stakeholders representing educators, students, parents, university faculty, foundation officers, business, industry and government officials, and senior citizens. The Collaborative's mission reflects its goal to prepare students for the 21st century through coordinating efforts and focusing resources to develop a regional approach to math and science education.

The region was originally defined as Allegheny County, the extended metropolitan area surrounding Pittsburgh, Pennsylvania, though the Collaborative extended its involvement to include the entire southwestern region of Pennsylvania during its second year of operation. Representation includes over 115 school districts, the Catholic Dioceses of Pittsburgh and Greensburg (á neighboring county's seat), and numerous private schools. Districts and schools represent a cross-section of urban, suburban, and rural institutions.

The Collaborative was formed in response to national and local reports of American children lagging behind their international counterparts in science and math achievement. A regional survey of parents sponsored by Miles (now Bayer), Inc. indicated that 95% believed that more children should be guided towards advanced science and mathematics coursework (Regional Math/Science Collaborative, 1997, p. 5). This, coupled with an emerging economy in technology-rich service provision, completing a transition from a strong, but declining, steel and manufacturing base in western Pennsylvania, clearly pointed to the need for regional planning and response in math and science education.

In 1994, the Allegheny Policy Council, a group of concerned business, industry and civic leaders, surveyed local school districts and found that schools were often acting in isolation on the same math and science priorities (Allegheny Policy Council, 1994). Additionally, while there were approximately 50 initiatives to strengthen math and

science education sponsored by universities, businesses and not-for-profit organizations, too few schools were aware of them, and few initiatives were aware of each other. These findings supported the need to prioritize and share resources and engage in a more focused, regional planning effort. An initial Stakeholder Congress was convened in March, 1994. The Congress recommended the establishment of the RMSC, to be located in the then-recently built Carnegie Science Center (an affiliate of The Carnegie, including the Carnegie Museums and Library, and a central location within the region), and guided by a Steering Council that would include representation from all concerned stakeholders. The Congress identified the three major priorities for regionally-focused mathematics and science education initiatives: infusion of appropriate technology to support instruction, development and use of instructional materials and assessments aligned with national standards for math and science education excellence, and the development and provision of quality professional development opportunities for area educators. As the Collaborative established an agenda of activities to address these priorities, data-driven decision making was considered integral to the planning process. Early planning documents reflect a critical need for data to drive the future planning and implementation efforts of the Collaborative (Bunt, 1998).

A focus on critical need, or crisis, can increase the perceived value of and the potential utilization of evaluation findings (Madaus, et al., 1983). The perceived crisis of falling achievement in math and science, coupled with a lack of clearly supported alternative solutions, provided fertile ground for the inclusion of evaluation data in policy analysis and revision. As a result, in 1994, the RMSC sponsored an "Inventory of Schools" survey to begin the process of clarifying school-based strengths and needs. This project yielded more questions than answers, and a decision was made by the Collaborative's Steering Council to seek a contract for external evaluation that would encompass an array of evaluation approaches and strategies to assist continuing regional discourse regarding math and science education. Stakeholder participation and supportive funding for the Collaborative was predicated on the ability to develop an efficient regional plan of action with stakeholder ownership, and to document its process and potential impact. The

prioritization of the Collaborative's mission and data-driven discourse, resulting in a regional plan of action, was crucial to sustain active participation of stakeholder groups.

The RMSC relies on intensive voluntary stakeholder involvement, coupled with only two paid full-time staff members and three consultants (including the evaluator). The Steering Council includes approximately 25 elected representatives, with each stakeholder group having equitable representation. The Council determines planning priorities, and charges the Managing Director to draft initial plans to address the priorities. The full Council or special subcommittees review the original draft plans, engage in collaborative dialogue, and refine the actual activities or initiatives subsequently offered for consideration to stakeholders throughout the region.

The Collaborative acts as a "broker," connecting stakeholders with one another through strategies and processes that allow shared ideas, dialogue, and resources among interested parties. Teachers across grade levels, disciplines and school-district boundaries share their concerns and pedagogical wisdom; professional development providers are able to hear directly from educators to inform their program offerings, and groups join together to examine issues related to how best to incorporate current research and curricular advancements in current and future practice. Information, in a variety of forms, both seeds and feeds these continuing processes, and evaluation plays a crucial role in the overall process.

Regional Math/Science Collaborative: The Role of Evaluation and the Evaluator

As a result of the need for continuing and expanded sources of information to facilitate and focus discourse, the Managing Director contacted the University of Pittsburgh with a proposal that provided the Steering Council's priorities, coupled with potential measurable indicators that could provide a description of the status of regional math and science education. The proposal requested an evaluation approach that would develop a useable "database" to provide meaning and context for the adopted indicators and support

continued dialogue and regional planning. Because of the priority for Steering Council participation, the Collaborative review process favored local evaluation consultants who could assure continuing contact and fuller participation of stakeholders within the process. The evaluation plan proposed through the University of Pittsburgh's School of Education was endorsed by the RMSC, and formal evaluation activities began in early 1996. The decision to use an external evaluator can support the development of a long-term relationship between the evaluator and the Collaborative, at a significantly lower cost than an internal evaluator (Mathison, 1994). Of additional benefit to the Collaborative, in light of future funding requests and high public visibility, an external evaluator may be perceived as more objective than internal staff, and may increase the credibility of evaluation findings across stakeholder groups and external reviewers (Mathison, 1994). By using an external contract the Collaborative is able to maximize evaluation potential and concurrently, minimize costs.

From the outset of the evaluation contract, the Collaborative was focused and clear regarding its information needs. Originally, the Steering Council developed a list of priority areas and began to clarify specific indicators that would link data to these priorities. Dialogue with stakeholder representatives examined the linkage of the data, representing regional "Indicators of Progress" (See Appendix A) related to math and science education, to district and regional planning. Through these efforts, the Collaborative had put in place an infrastructure to clarify evaluability and encourage utilization of subsequent evaluation findings. The resulting evaluation plan sought to collect and clarify baseline data regarding the status of math and science education indicators in the region, with additional plans to include annual updates to the planning database. Planning followed a similar strategy within the Collaborative, with a primary focus on establishing an "inventory" of what was currently available or established, prior to inviting stakeholder dialogue to considering alternatives and priorities for shared resources. The linkage between evaluation and planning was deeply integrated within the processes of discourse and regional planning through regular and continued dialogue among the evaluator, the RMSC staff, and the Steering Council. Additional opportunities

BEST COPY AVAILABLE

to share information with all stakeholders included publishing summary reports in an annual publication delivered to over 40,000 area math and science teachers, and the inclusion of information sessions and utilization workshops at semi-annual meetings attended by over 400 stakeholders. The linked developmental process of planning and evaluation strengthened the bonds between and among stakeholder groups through opportunities to deliberate about priorities, strategies, activities, and the means to continue gathering information.

This *a priori* emphasis on formally linking evaluative inquiry within the planning process and continuing review of priorities and strategies helped to focus a clear vision for the development and subsequent changes to the evaluation plan. The RMSC, through the elected Steering Council, and the Director, became active partners in the development of the evaluation plan. An inherent feature of participatory evaluation models is the continuing contact and discourse between the stakeholders and evaluator (Cousins & Earl, 1995). By continuous representation of stakeholder concerns, the Steering Council assures that stakeholder needs are explicitly reflected and addressed within the planning and evaluation efforts. Through this participatory model, the match between the evaluation activities and the priorities and needs of the Collaborative, and each stakeholder group, is continually reviewed and adjusted. With each review cycle, stakeholder affiliation with the Collaborative is affirmed. The close linkage of planning and evaluation within the participatory model, helps to build and sustain relationships between and among stakeholders while concurrently focusing and informing the discursive practices of the Collaborative in determining regional priorities and planning processes.

Patton indicates that the nature of the interpersonal relationship between the evaluator and the stakeholders has "substantial implications for the use of program evaluation." (Patton, 1986, p. 45) He argues that the presence of a clearly identified individual or group of stakeholders who care about the evaluation is essential to utilization. The Steering Council and the management team of the RMSC act as this interested and active

group. The Steering Council, via its original emphasis on evaluation linked with planning and development, and the insistence of stakeholder representatives to be included in each step of evaluation planning, assures continuous involvement and use of evaluation findings. As mentioned previously, this involvement not only provides a continual check-and-balance process for the evaluation plan and activities, but also serves to reinforce stakeholder affiliation and engagement. The Collaborative exhibits a strong commitment and responsiveness to evaluation, and evaluation is an integral component to the continuing dialogue and review of priorities and options for action.

Stakeholder groups include teachers and school administrators, groups often disillusioned with the promise of evaluation; business and industry representatives, who are focused on the "bottom-line" of effectiveness and efficiency; and officers of foundations that partially fund the RMSC, disenchanted with process-evaluation and hungry for outcome and impact measures. The Collaborative infrastructure supports and encourages ongoing participation of Steering Council members and staff to build and maintain credibility of evaluation efforts among diverse stakeholders. As a result, development and refinement of the evaluation plan, the Indicators of Progress, and resulting instrumentation and other evaluative strategies was specifically reviewed, and in some cases, revised by RMSC staff and Steering Council members. The evaluation plan became focused on measurable indicators of change in the regional status of math and science education. For example, early discussion of "attitude shifts" among teachers and students were summarily dismissed as indicators of progress, in favor of measurable impact on teaching practice and student achievement. Customary tracking of enrollment patterns in upper-level math and science courses was deemed less important than documented *successful completion* of these courses by high-school graduates. The Steering Council was clear in its focus on meaningful, documentable impact, and equally clear about their commitment to ongoing involvement in both the Collaborative's wider, as well as the more evaluation-specific, dialogue.

The evaluator meets on a quarterly basis with the full Steering Council. The Steering

Council reviews drafts of evaluation reports to suggest modifications and recommend avenues for appropriate dissemination of findings. The evaluator is perceived by stakeholders as an integral part of the Collaborative "team," an extension of the RMSC staff. This ongoing and expansive relationship enhances the partnership between the evaluator and Collaborative. The ongoing participation of stakeholders with the evaluator, and in the evaluation process, further affirms the value of stakeholders within the collaborative effort.

Perceived, and treated as an extended staff member, the external evaluator has gained the advantages of an internal evaluator: access to information and organizational culture, a fuller perception of operating limitations and potentials, and a richer understanding of the context for the evaluation. Within this structure, the boundaries of the relationship must be constantly renegotiated: How often, and to what extent, will the external consultant participate in internal activities? How will the evaluator and the organization describe the relationship they share? What pressures are brought to the evaluation plan and the evaluator, from both external and internal sources? The relationship between the evaluator and the Collaborative succeeds only to the extent that each partner is willing to authentically engage in the dialogue, and the resolution of potential conflicts, within these guidelines.

As the evaluator for the RMSC, I believe my professional history as an evaluator with key members of the Steering Council and the Managing Director has facilitated a relationship of mutual trust and professionalism to flourish. The Managing Director had served on an advisory board, and some key stakeholder representatives had been associated with a longstanding public school project for which I had served as an internal evaluator in prior years. I had, in effect, established my initial credibility via past acts. Additionally, through an eclectic participatory approach to the evaluation, remaining flexible and open to changing priorities and matching evaluative techniques to the evolving needs of stakeholders, the Steering Council perceives the evaluation process as a working blueprint, open to modification via their direct participation. Utilization of

evaluation findings in strategic and operational decision-making would surely suffer with a more stringent, less responsive approach within this collaborative context.

Lessons Learned and Implications for Further Study

Collaborative educational initiatives present opportunity and challenge for evaluation. The complexities of diverse stakeholder perspectives require a sensitivity to the organizational culture of the collaborative for successful implementation and utilization of evaluation studies. The "context factors" related to evaluation utilization delineated by Cousins and Leithwood (1986), including the information needs, political climate, decision characteristics, commitment and responsiveness to the evaluation, and other "personal" factors (Patton, 1986), determine the initial, and subsequent, parameters of the relationship between the collaborative initiative and the evaluator. Lack of consensus among stakeholders about the purpose or the intended uses for the evaluation (Barabba & Zaltman, 1991) and the lack of participation by key decision-makers in the overall evaluation processes (Alkin, Burry, & Ruskus, 1984; Cousins & Earl, 1995) may contribute to valuable evaluation findings left under-utilized within the organization. As an external consultant, the evaluator operates from a contextually-removed locus, and must work diligently to develop a broad understanding of the organizational and interpersonal relationships that are represented in the collaboration, while fostering fuller participation of stakeholders. Rarely, do two-dimensional representations such as organizational charts and summary documents, reveal the intricacies of interpersonal nuance which operate among stakeholder, staff, and the evaluator. To be effective, the evaluator must gain and develop extended credibility with stakeholders to build and maintain both a deep, and broad organizational understanding and contextual sensitivity. Ongoing communication with stakeholders, spending more time listening than speaking, and responding directly to stakeholder needs that have been expressed, serves to develop credibility and a sense of cooperation. Stakeholder "ownership" of the evaluation process is crucial for sustained cooperation. Building sustainable relationships, which serve to prioritize the evaluation process within the organization, requires a large investment of time and energy, coupled with a willingness to be simultaneously flexible

and focused in the definition of the evaluator's role. The evaluation plan, and the role of the evaluator, must remain flexible and responsive to evolving needs. Through the development of credible partnerships with staff and stakeholders, the external consultant can achieve the advantages normally associated with an internal evaluator, including access to information and stakeholders, a fuller understanding of the organization and its culture, and gain stakeholder and staff cooperation necessary for extensive data collection and use.

To best serve diverse stakeholder concerns, I attempt to enter the dialogue adequately informed and technically prepared with evaluation approaches and strategies, yet I remain open to the diverse needs of the collaborative *and* the specific needs of stakeholders. Utilization, like the evaluation planning process, has taken on a variety of forms and requires diverse strategies and ongoing negotiation. Utilization of evaluation information is cultivated through relationships between and among stakeholders, as well as between myself and stakeholders. The RMSC relies on a strong foundation of need for useful data as a starting point for stakeholder discourse regarding priorities and program-level decisions. As the external project evaluator, I have often been called upon to present evaluation data from the perspective of an interested stakeholder, rather than from the position of a neutral, objective external consultant. The evaluation data are presented at the table for continuing dialogue and reflection, in a manner similar to other information brought forth by other stakeholders. Through this approach, evaluation data may at times begin the dialogue, may become a clarifying mechanism within an existing discussion, or may imply the need for additional information prior to continued discussion or decision points.

Within this collaborative initiative, I am viewed as an extension of the Collaborative staff, or as a stakeholder representative rather than solely as an outside consultant. I represent the perspective of critical review and evaluative technique, coupled with a deep concern for the context of the organization and its mission. Professional boundaries can easily blur. Individual credibility, and professional guiding principles and standards for

evaluation can provide both the stability and flexibility to allow the consultative relationship to expand, particularly in terms of the stakeholders' information needs and the methods for gathering and reporting relevant information to substantively add to regional discourse to strengthen math and science education opportunities.

Collaborative educational initiatives hold promise to impact the quality of education in crucial areas. They facilitate and broker a mix of resources, both human and financial, that targeted efforts can rarely garner. Evaluation can play a critical role in examining the potential of these efforts, and simultaneously, help to build and focus the partnerships in ways that facilitate their success. This requires an expanded definition of evaluation and evaluator roles, diverse models of approach, and strategies of interaction and negotiation to better fit the diversity collaborative organizations reflect. There is a need to balance and juggle stakeholder needs and concerns. Within this context, I have become both a participant and a facilitator of dialogue to assist in building a deeper and richer understanding among stakeholders. Preconceived evaluation approaches have been put aside to be more responsive to the clarified evaluation needs of the organization that percolate through discursive practice. Once determined, these needs drive the further development and implementation of the evaluation plan. As the evaluation and my involvement have matured along with this Collaborative, new needs have arisen. Currently, we are exploring a variety of strategies designed to assist schools and districts, alone and in partnership with others, to build internal evaluation capacity within their home organizations. This further expands the role I play in educative and consulting arenas, and holds even more challenge and benefit to increase the value for and utilization of evaluation.

As I have come to understand the dynamics of working with the RMSC initiative, I remain willing to further reflect on both the role of evaluation and my role within the Collaborative, and to more fully examine how evaluation can best serve the Collaborative's efforts. As a concerned and reflective practitioner, I attempt to refine my ability to provide meaningful information to stakeholders, and through participative

evaluation, assist in the Collaborative's continued discourse related to science and math education.

The issues inherent in participative approaches to evaluation, especially in a diverse collaborative setting like the RMSC, are complex and challenging. To more fully explore these issues and better frame the focus of continued study of evaluation and evaluator roles, it is important to review relevant literature. The historical and philosophical underpinnings of past and current evaluation practice can not only illuminate the important issues and changes in over time, but can point to evaluation conceptions and approaches that may prove most beneficial in this type of setting.

History and Traditions of Program Evaluation: Brief Review

While formal program evaluation in education was virtually non-existent prior to the mid-1800's, early examples of the use of information to draw conclusions, make decisions, inform choices or judge the value of a person or program do exist. Early Chinese officials, approximately 2000 BC, used civil service examinations, administered every three years, to evaluate worker competency. Workers who passed were retained and promoted, those who failed were summarily dismissed (Travers, 1983). Though additional early examples may be found, prior to the mid-1800's, religious and political beliefs determined the outcome of most educational issues and the need for additional information to inform choices was minimal.

The state education departments of both Massachusetts and Connecticut were instrumental in establishing early strategies to collect information for educational planning purposes. Between 1838 and 1850, Horace Mann and his colleagues submitted twelve annual reports identifying and exploring educational issues and concerns of the era and included the use of empirical data to support their claims. These early attempts were designed to influence policy decisions at the state level and continue, to present, to serve as the foundation for evaluation in state and federal education authorities (Worthen

& Sanders, 1987). In 1845, the Boston School Committee commissioned what came to be known as the Boston Survey. This evaluation constituted the first recorded use of printed tests for the broad assessment of student achievement. A sample of Boston's schoolchildren was tested in all major areas of the curriculum including geography, grammar and definitions, history, philosophy, writing and arithmetic. Test results were accumulated over two years, 1845 and 1846. While the School Committee was shocked by the low level of achievement revealed by the test scores, the test was abandoned because of a lack of utilization of results to change pedagogy or improve student learning (Travers, 1983).

Joseph Rice conducted a similar project across numerous urban school systems in the country between 1895 and 1905. Rice was a vocal critic of the educational practices of the era, and was highly motivated to support his claim of inefficient use of school time through the presentation of empirical data. He reported minimal differences in achievement in spelling across schools, regardless of techniques employed to teach and practice spelling. The study also revealed substantial differences in arithmetic achievement, and he used these results to propose the need for the development of a standardized test of these skills, to more accurately detect and describe the differences (Travers, 1983). Rice also may have been the first proponent of the judicial, or advocate/adversary model of evaluation. He proposed that controversial issues or decisions might best be resolved by gathering relevant data, both pro and con, and presenting the data to a qualified panel of judges for an impartial hearing to determine the outcome (Rice, 1915).

As a result of the efforts of Edward Thorndike, the father of educational testing, significant increases in evaluation activity occurred during the early 1900's (Worthen & Sanders, 1987). Thorndike persuaded the educational community that refined and developed testing techniques were essential to accurately measure student abilities and achievement. By the 1920's many large school systems had established bureaus of school testing that were charged with managing and implementing large-scale

assessments of student achievement and curriculum effectiveness (Worthen & Sanders, 1987).

During this same period, the Superintendent of the Gary, Indiana public schools commissioned an evaluation report to substantiate his claims that Gary's students were among the best in the country. The final report provided evidence quite to the contrary, suggesting that Gary students were less able than a comparison cohort, though some later observers indicated that the study design was biased against the Gary curriculum strategies (Worthen & Sanders, 1987).

Stufflebeam and colleagues (Madaus, Scriven & Stufflebeam, 1983), acknowledge the broad, and lasting, influence of Ralph Tyler across decades of educational evaluation, by referring to the period of 1930-45 as the "Tylerian Age." Ralph Tyler, employed in the early 1930's as the study director for the landmark Eight Year Study, sought to compare progressive Deweyian curricula and the more traditional objectives-oriented Carnegie-unit curricula in relation to pre-college preparation and entrance. Tyler and his colleagues developed a number of instruments designed to measure performance on educational objectives. Their work was premised on the direct linkage between stated objectives and achieved results. The work of Tyler (1942) dominated evaluation discourse for many years, and can still be identified as the underlying logic in many evaluation designs today (Worthen & Sanders, 1987). Tyler continued to be an active evaluator and was the original designer of the National Assessment of Educational Progress (Travers, 1983).

During the 1930's, educational accreditation agencies flourished and gained credibility and power (Stufflebeam, 1969). Periodic accreditation reviews replaced the rather burdensome school inspection systems in place prior to this period. Contrary to the Tylerian model, the accreditation process examined variables related to the capacity of a system to provide quality education, for example the availability and adequacy of human and financial resources. Accreditation efforts represented the first widespread

institutionalization of school evaluation across U.S. public schools, and served as the impetus for the development of process and feasibility evaluation models (Worthen & Sanders, 1987).

The period of 1940 through the 1960's was marked by technical refinement and consolidation in evaluation. Advances in the field included incremental fine-tuning of instrumentation and technique rather than deep or sweeping reform of evaluative practice or underlying assumptions. Evaluation efforts continued to be influenced by positivism and this approach was strengthened by the development of taxonomies of educational objectives by committees chaired by Bloom and Krathwohl (Bloom, et al., 1956; Krathwohl, et al., 1964). Program and student evaluation was centered squarely on monitoring outcomes in relation to behavioral objectives, a natural extension of Tyler's approach. No other models or approaches gained prominence during this period.

The launch of the Russian spacecraft, Sputnik, resulted in swift action and legislation impacting educational policy and programs across the United States. The National Defense Education Act of 1958 provided development funds for renewed curricula, especially in math and science education. Subsequently, funds were appropriated to begin evaluation efforts linked to these new curriculum options (Worthen & Sanders, 1987). The greatly expanding need for quality evaluations was beyond the existing resources of evaluators to respond. Cronbach (1963), in a seminal reflection on the quality of evaluations of the era, reported that most evaluations of the period were less than helpful to the burgeoning task of education, and called for new directions in educational evaluation. He claimed that more information was needed to provide insight about the programs *as they were implemented to augment improvement efforts*, rather than relying on a review once the program was well-established or completed. Later, Scriven (1967) distinguished two types of evaluation, formative and summative, that would become universally accepted concepts among evaluators. Formative evaluation calls for a thorough review of the program *while the program operates for the purpose of improvement*, while summative evaluation looks at the worth or merit of the program at

the *conclusion* of its activities and initiatives based on outcomes or effects. This definition is rather oversimplified, and often the distinction between formative and summative evaluation blurs in actual fieldwork (Stake, 1969), however, both Cronbach and Scriven were acknowledging the same need: evaluation of educational programs during implementation, for improvement. The Cronbach article generated a new energy within the circle of professional evaluators, and stimulated an increased discourse of evaluation purpose and practice (Worthen & Sanders, 1987).

In addition to these discourses, additional legislative efforts to impact educational outcomes and the resulting discussions of school-based intervention also helped to shape evaluation practice. As the civil rights movement sprang into the foreground of political discourse, and the Coleman Study (*Equality of Educational Opportunity*) of 1965-1966 was commissioned as a result of the Civil Rights Act of 1964. The Coleman Study was a large-scale policy evaluation of issues of racial inequity and achievement commissioned by Congress. In the report, Coleman concluded that there was no evidence of racial inequality based on resources available. He further stated that "schools bring little influence to bear on a child's achievement that is independent of his background and general social context" (Coleman, Campbell, Hobson, et al., 1966, p. 325). In part, as a counterargument to the Coleman Study, a genre of educational research, known in retrospect as the "effective schools" literature, flourished in the 1980's. Heralded by Ron Edmonds (1979), the literature makes the case that indeed schools can and do make a difference, beyond the limited factors identified by Coleman. School evaluations were directly designed to measure "effectiveness" based on the five correlates of successful schools outlined by Edmonds (a safe and orderly climate, a common sense of purpose and mission, strong instructional leadership, high expectations for students, and frequent monitoring of student achievement data). The literature was further expanded to include the addition of an emphasis on instructional time (Fisher, et al, 1980) and parental and community involvement (Comer, 1980). Within this framework, a school was defined as effective "when it brings low income children to the minimum basic skills mastery level which now describes minimally successful performance for middle income children"

(Shoemaker, 1986). Numerous state departments of education, led by efforts in Connecticut, developed extensive evaluation and research bureaus and support systems based on this design.

Another result of the civil rights movement, and the one of the most significant legislative influences on educational evaluation, was the enactment of the Elementary and Secondary Education Act (ESEA) of 1965. The act authorized a variety of educational initiatives, however, the most far reaching and largest appropriation was in Title I of the Act (later known as Chapter I), which earmarked funds for educational intervention strategies targeted for disadvantaged youth. The ESEA unleashed millions of dollars through tens of thousands of federal grants to local, state, and university-related educational agencies. Senator Robert F. Kennedy was the most vocal advocate for including a requirement for annual achievement testing within each grant, a far-reaching attempt to hold agencies accountable for documenting effects for the vast dollar amounts they would receive (Worthen & Sanders, 1987).

Once again, public education could not adequately respond to the pressing need for evaluation. Very few districts employed specialized evaluation personnel, and often, the district would appoint a teacher to serve as the internal Title I evaluator overseeing the mandated annual testing. The United States Office of Education (USOE) never fully operationalized Congress' guidelines into specific recommendations related to evaluation. The resulting evaluations were generally of limited value because of the variety of testing procedures, the lack of standardized test instruments and administration, and a high number of absentees from test sessions. The lack of information or assistance, either from Congress, or the USOE was not for lack of effort, but rather, lack of knowledge, expertise, and conditions conducive to high-quality evaluations (Worthen & Sanders, 1987). Prior to this period, educational evaluation had focused on the development and use of student achievement testing, and little theoretical work was in place to support the need for expanded program evaluation, especially in-process, formative reporting of program effectiveness called for by Cronbach and Scriven.

The growing demand for evaluation of educational programs and other initiatives, coupled with the emerging discourses regarding alternative approaches helped to establish evaluation as a young and flourishing profession. Leaders emerged from among scholars and practitioners representing a variety of disciplines and many evaluation models and approaches were developed and implemented. House (1980, 1983) has offered extensive reviews of the philosophies of "knowing" and competing world views that have influenced educational evaluation. Paradigmatic diversity among evaluators has served to both expand the discourses available to practicing evaluators, and at times, polarized practice through the rigid alignment of method with a particular world-view. As discourses in other disciplines such as sociology, psychology, economics, and others informed each other, so too, their influence was felt in evaluation circles.

The growing profession was further strengthened, through the late 1960's and early 1970's, via the encouragement of education-related professional associations, including the American Educational Research Association (AERA) and the Association for Supervision and Curriculum Development (ASCD) pressing members to attend more directly to evaluation needs. AERA established Division H for school-based evaluators in 1971, and in 1975 Phi Delta Kappa provided seed money to facilitate the formation of The Evaluation Network, the first professional association exclusively for evaluators. As the Network grew, the group published *Evaluation News*. A complementary organization, The Evaluation Research Society, was formed in 1976. In 1986, the two sister professional organizations merged, forming the American Evaluation Association (AEA), which now publishes a variety of materials including the refereed journal, *American Journal of Evaluation* (formerly *Evaluation Practice*) and sponsors an annual national conference (Worthen & Sanders, 1987). Additionally, the profession joined with representatives of major educational stakeholder groups to form the Joint Committee on Standards for Educational Evaluation. Since their inception in 1975, the Joint Committee has developed and revised standards of practice related to program evaluation in 1981 and personnel evaluation in 1988 and published updated standards and case study reviews

(Joint Committee on Standards for Educational Evaluation, 1994). Additionally, the AEA has developed a general set of "guiding principles" to inform both the membership and potential education "clients" of the basic tenets of good program evaluation practice (Shadish, et al., 1995).

As the profession was informed by a variety of disciplines and applied evaluation practice, a variety of models, metaphors and conceptions of educational evaluation were generated. Early definitions equated educational evaluation with student achievement testing and measurement. Accreditation site visits, still a popular form of both internal and external review, have also been characterized as educational evaluation. Tyler's influence on evaluation led to a strong emphasis on objective or goal attainment as a predominant conception of educational evaluation. Scriven (1967) focuses on the evaluative role of summative judgement to determine the merit or worth of a program or activity, coupled with formative evaluation methods to guide program improvement. Various purposes of evaluation have also influenced conceptualization of evaluation, including a focus on evaluation utilization. Patton (1986) offers the definition "systematic collection of information ... to make judgements about the program, improve program effectiveness, and/or inform decisions about future programming." He further clarifies the role of utilization-focused evaluation as evaluation activity "done for and with specific, intended primary users for specific, intended uses" (Patton, 1997, p. 23). Others specify the role of evaluation is "to help educators as they consider issues surrounding educational policy, as they establish priorities for improving educational systems, or as they engage in the day-to-day management of educational systems" (Cooley & Bickel, 1985, p. 3) or "delineating, obtaining, and providing useful information for judging decision alternatives" (Stufflebeam, 1973, p. 129).

Diverse conceptions and definitions of evaluation have led to the development of a variety of approaches and models. Philosophical and definitional characteristics influence how evaluation is conducted and the role of both evaluation and the evaluator within program settings.

Growing Profession: Early Models and Conceptions of Evaluation

Educational issues worthy of the scrutiny and the discernment evaluation might offer are many and varied. In an attempt to better understand the intent and potential usefulness of evaluation models, a variety of authors have published comprehensive categorizations and reviews of the many models developed during the 1960's, 70's and 80's. Authors cluster models by the methods employed (Talmage, 1982), by the epistemological and ontological perspectives that shaped the design (House, 1983, 1991, 1993), still others by the context and specific needs of the evaluation (Worthen & Sanders, 1987). For the purposes of this paper, a discussion of the changes in the conceptual roles of evaluation and the evaluator are crucial to more fully informing the study of evaluation practice within the RMSC.

Most of the early formal educational evaluation models were variations of the Tylerian objectivist approach and relied heavily on a positivist world-view and rational science methodology. Based on a research design to create generalizable knowledge, evaluators were hopeful that well-designed models might offer specific information to inform individual projects and also contribute to a wider applicability through comparison and cross-site application of the same, or very similar, models (Cronbach & Suppes, 1969). Evaluation models often suggested a "template" approach that could be applied with minor adaptation across programs and projects, and evaluation teams and individuals were often associated with a particular model that was applied in much of their evaluation work, such as, Dan Stufflebeam and the C.I.P.P. (Context, Input, Process, Product) model (Stufflebeam, 1971, 1983) and Mal Provus and the Discrepancy model (Worthen & Sanders, 1987).

Most evaluation models sought to offer a summative review of the program, focused on measuring outcomes based on pre-determined variables. Evaluators used techniques, often borrowed from prevalent positivist research methods, that allowed for an objective

review of stated program inputs and quantifiable outputs, striving to establish causality within programs that often neglected to articulate a program theory of action that would support this claim. As the number of models increased, so did the variation in application, and suggested approaches that used a "mixed methods" or eclectic set of techniques and multiple sources of data were discussed and critiqued among evaluators (Patton, 1986; Cooley & Bickel, 1985; Scriven, 1984; Worthen & Sanders, 1987). These approaches relied more heavily on formative data to better describe and value the program's process and facilitated client's mid-course adjustments to implementation. Additionally, there was a growing recognition of multiple perceptions of program features and parameters across various stakeholder groups.

It is fortunate that many of the early pioneers in the educational evaluation profession are still active contributors and discussants. The American Evaluation Association (AEA), the primary professional association for U.S. evaluators, sponsors an online discussion forum for its members and other interested evaluators. Over 1380 evaluators worldwide subscribe to the "EVALTALK" listserv. As a way to augment the literature reviewed for this paper, I presented a series of questions for subscribers, specifically asking senior colleagues, as well as others in the field, to comment.

I asked colleagues:

> "Was the intent of the early models originally to provide the "best" model for program evaluation; one that could be replicated and applied across settings?"
> and,
> "Were later models an attempt to refine existing models' perceived inadequacies? as a refinement? as a response from a different philosophical perspective?"
> (Tananis, 1998)

In response, William Shadish, an active contributor and past-president of AEA, indicated that "each person truly believed the model they proposed would be a positive development in the field, and would be useful in contexts other than their own" but

cautions that "[early developers] were probably too smart to think it would be "best" in a field as new as evaluation." He further represents the focus of the profession on evaluation *practice* during the 1960's through the 80's indicating "they were doing their best to advocate their own model and didn't really spend much time doing "comparative evaluation theory to show their theory was best" (Shadish, 1998). Evaluation discourse during this period was focused on refining and expanding practice, much more so than explicating or investigating issues of theoretical perspective that cut across practice.

Robert Stake, a longtime evaluator, often associated with his work in case study research and evaluation, adds that the "emergence of models was driven by disappointment with existing practice," specifically referring to previous "goal-oriented empirical studies" borrowed directly from scientific research. Stake admits that "there was some modesty about the power and appropriateness of the models, but we authors and colleagues saw a competition among them and felt that within certain constraints, our particular advice was better than other advice" (Stake, 1998). While perhaps the developers of early formal models did not necessarily intend that their models become formal systems adopted in whole, or even part, across a variety of settings, the overall context and need for evaluation during this active period may have set the scene for abuses of the limitations of the models in practice. Stake belies this problem as he reviews his own contribution by saying, "my own writing in 1967 was called by some "the countenance model" but was [intended] only [as] a categorization of potential data, not a guide for carrying out a study" (Stake, 1998).

His experience is mirrored by Michael Scriven, a prominent evaluator often associated with consumer-driven evaluation, who reflects that "goal-free evaluation was widely assumed to be a model I was proposing for all of evaluation; it was never more than an approach to be used where appropriate" though he makes the claim, in the same reflection, that Stufflbeam's CIPP (Context, Input, Process, Product) model was intended to be used across all program evaluation contexts (Scriven, 1998). Michael Quinn Patton, summarizes many of the responses by saying "pluralism was an ethic from the beginning,

in part because evaluation was interdisciplinary from the beginning; just as no discipline could/would dominate, no model would/could either" (Patton, 1998). Scriven captured the essence of the increasing demands for the development of high-quality approaches to evaluation by commenting that "the proliferation of evaluation models [was] a sign of the ferment of the field and the seriousness of the methodological problems which evaluation encounters. In this sense, it [was] a hopeful sign" (Scriven, 1984, p. 49). Various approaches and conceptions of evaluation were generated in response to the diverse informational needs within specific educational contexts, coupled with the inability of experimental or "borrowed" research methods to fully meet these needs.

While individual reflections on evaluation history make the case for "pluralism" (Patton, 1998), the "competition" alluded to by Stake (1998) and Shadish (1998), and the influence of particular models or approaches by various education agencies which then drove practice, argues for a less eclectic early period than may have been intended by the evaluation theorists who first published the works cited by the agencies who applied them. As noted earlier, the dire need for evaluation strategies to document program effectiveness and accountability, and the use of evaluation strategies by practitioners not necessarily immersed in evaluation theories, encouraged a utilitarian adoption of models or approaches, perhaps extending beyond the original intent of the architects.

Within this same period, critiques of existing evaluation practice created an expanded discourse. House noted a shift in philosophical perspective that informed evaluation practice. In his review of prevalent evaluation models based on an "objectivist" world view, he points out that these models relied on a rational scientific perspective with claims of validity and reliability supported through rigorous statistical analysis of quantitative data. The models relied on their methodology for credibility and validity, and were generally inattentive to competing world views that held "objectivism" as limited in the types and usefulness of information yielded (House, 1983). House published numerous critiques of then-current evaluation discourse, pointing out the need for evaluation to be more sensitive to the perspectives of both critical theory and

interpretive thought (House, 1980, 1983, 1991, 1993). Campbell mirrors House's sentiments by adding "20 years ago [1960's] logical positivism dominated the policy of science ... today the tide has completely turned among the theorists of science in philosophy, sociology, and elsewhere" (Campbell, 1984, p. 27). Shadish (1998), reviews some of this early history by reminding us of a broader context for paradigmatic perspectives. He cautions that "... early evaluators came from other fields entirely, bringing their preferred methods with them, only to encounter issues specific to evaluation a bit later as they tried those models and methods out in the evaluation context." As evaluators struggled with issues specific to the evaluation context, such as stakeholder involvement, competing perspectives and purposes for evaluation, and issues of relevance and meaning to support fuller use of both the evaluative process and the findings, the discourse, at an increasing rate, focused on a set of emerging philosophical and methodological issues across the field rather than issues specific to individual models or approaches.

Guba and Lincoln propose that we are experiencing a paradigm shift from the positivist influences of rational science to a constructivist approach in research and evaluation. The failure of positivist-generated models to adequately reflect and clarify the complex realities present in educational endeavors led to approaches that were based on non-scientific metaphors, including judicial adversarial models (Rice, 1915), responsive approaches (Stake, 1975) and artistic connoisseurship (Eisner, 1976). Guba, in 1978, offered a "naturalistic" alternative that is characterized by a relativist ontology, accepting multiple-socially-constructed realities, a subjectivist epistemology, viewing the researcher and researched constructing the inquiry jointly, and a hermeneutic methodology, embracing context as a crucial informant to the phenomenon under study (Guba, 1986). The resulting evaluative inquiry is a process of negotiated and renegotiated areas of focus, criteria for the evaluation, and the roles of the evaluation and the evaluator. Guba and Lincoln offered a set of criteria to judge naturalistic research and evaluation that paralleled the positivist concepts of reliability, validity, generalizability, and objectivity that included dependability, credibility, transferability, and confirmability,

respectively (Guba, 1981; Guba & Lincoln, 1981). Interpretivist critics suggested that these criteria mislead one to assume that "the two approaches [positivist and naturalist] are variations in techniques within the same assumptive framework to reach the same goals and solve the same problems" and were thus, inadequate (Smith & Heshusias, 1986, p. 6). In response, Guba and Lincoln offered a set of "criteria of authenticity" including fairness in representing multiple realities, ontological authenticity embracing context as central to understanding, and educative, catalytic, and technical authenticity that facilitates a deeper understanding which leads to action and empowerment (Guba & Lincoln, 1986). While supporting the use of mixed quantitative and qualitative methods in evaluation, Guba and Lincoln propose that differences in philosophical perspectives cannot be resolved, as Lincoln offers, "there can be no compromise, no integration" since each paradigm assumes a different world-view and subsequently, different research interests and mathods (Guba and Lincoln, 1989, p. 2). They propose that evaluators must be responsive to multiple realities rather than applying a preconceived or privileged singular reality. Existing discourses within the field were centered on the purposes and intents of evaluation (program development and improvement versus judgement) and the use of various qualitative and quantitative methods to support these purposes. Guba and Lincoln increased the stakes within the debate by proposing that paradigmatic perspective was central to definitional and methodological decisions. These issues are still hotly contested and debated at both national and international levels within educational research and evaluation fields.

Rethinking Educational Evaluation: Emerging Issues and Practice

The vast number of evaluations contracted and completed during the late 60's, 70's and 80's provided a considerable body of experience to inform future directions. Professional discussion and debate turned to issues *across* the application of approaches and specific models, rather than being focused on the development of particular models or strategies. The paradigmatic discussions in social research and other disciplines regarding epistemology and ontology continued to influence evaluation discourse. The diversity of

experience and formal training that evaluators brought to their "new" field of study and practice provided fertile ground for broad-based discourse. Discussion of paradigmatic perspective, stakeholder involvement, utilization and related issues emerged as practitioner-scholars of the era reflected on the decades of practice, as Shadish reminds us "dealing with [these issues] required more experience with evaluation than early evaluators had." (Shadish, 1998)

As discourse in evaluation focused on issues of fairness and representation of multiple perspectives coupled with concern for increasing evaluation utilization, a variety of participative approaches to designing and conducting evaluations emerged. Stakeholder approaches sought to address "the lack of fit between evaluation and the sociopolitical context of the program world" (Weiss, 1983, p. 3). Evaluation was criticized as too narrow in its scope, limited in its consideration of indicators of success, and ultimately responsible to more powerful sponsors who commissioned the evaluation. All of these factors limited usefulness and utilization of evaluation findings by stakeholders (Weiss, 1983, p. 5).

Why Use a Stakeholder Approach?

Weiss has summarized many of the major themes that support the inclusion of stakeholders in the evaluation process:

> "The stakeholder concept represents an appreciation that each program affects many groups, which have divergent and even incompatible concerns. It realizes and legitimizes the diversity of interests at play in the program world. It recognizes the multiple perspectives that these interests bring to judgment and understanding. It takes evaluation down from the pedestal and places it in the midst of the fray. It aims to make evaluation a conveyor of information, not a deliverer of truth; an aid, not a judge" (Weiss, 1983, p. 11).

Murray elaborates on the potential benefits of stakeholder involvement, indicating that it is "a useful device for getting leading players to cooperate, for understanding a program

intimately, for attracting attention to interim evaluation findings, and perhaps even for getting decision makers to take evaluation findings into account when they make decisions" (Murray, 1983, p. 59).

Evaluators, working through regular consultation with various parties with interest in the evaluation, solicit multiple views related to all aspects of the evaluation, including design, ongoing modification, and ultimately, response to and use of the evaluation (Cohen, 1983, pp. 73-74). This stance represents a shift from the more distanced "objectivist" (Stufflebeam, 1994) evaluations resulting in an evaluator's judgement of "merit or worth" (Scriven, 1967) to a more developmental focus, or as Patton frames it, as responding to a "different type of evaluative question" (Patton, in Alkin, et al., 1990, p. 116).

Weiss (1998) indicates that a shift to participative forms of evaluation is sometimes driven by the evaluator being uncomfortable with the "power imbalance," where the evaluator stands in judgement of a program with limited input from the stakeholders most directly impacted by evaluation findings (p. 101). Contrary to the notion of objectivity and professional distance so entrenched in the positivist perspective, House denies that this "state of grace" ever really existed in evaluation or research, in general. Further, he suggests that such a perspective serves only to minimize, if not totally ignore, stakeholder needs and goals. He proposes that evaluators, in their attempt to remain objective, contrary to remaining extracted from constituent agendas, had fallen prey to hearing and valuing only the program manager's agendas, to the exclusion of competing or complimentary views (House, 1991, 1993). Rather than a sterile process of information gathering and analysis determined exclusively by explicit program goals, House recommends a more open process where evaluation methodologies are determined by program realities including stakeholder perspectives and goals, and planned as well as consequential, or unanticipated, outcomes. Within this framework, evaluators should be prepared to apply a wide variety of quantitative and qualitative techniques; and should actively seek and reflect a multitude of sometimes competing agendas and concerns.

Weiss, describing stakeholder evaluation, appears to support the infusion of values in evaluation:

> "Realization of the legitimacy of competing interests and multiplicity of perspectives and willingness to place evaluation at the service of diverse groups are important intellectual advances ... The concept enfranchises a diverse array of groups, each of which is to have a voice in the planning and conduct of studies" (Weiss, 1983, p. 11).

House argues that issues of diversity such as feminism and racism have been accommodated in most evaluation schemas, but that the issue of social class has remained widely ignored. He proffers that evaluators often assume that issues of the poor and powerless have been incorporated into programs explicitly, hence there is no need for the evaluator to further question the inclusion, or potential exclusion. He notes that most professional evaluators themselves, are members of a more elite and powerful social class, and that their own bias, based on personal social class, acts to support this often erroneous notion. In more recent writings, House has identified what he terms "ethical fallacies" (House, 1993, p. 168). House claims that evaluators often rely on clientism (I am evaluating what the *client* wants), managerialism (My audience *is* program managers), methodologicalism (I am, methodologically, performing the *"right"* techniques), relativism (Everyone's input has *equal* weight), and elite pluralism (Diverse inputs have been negotiated and are *adequately expressed* by the elite) to justify their evaluation design and conduct as ethical. These "fallacies," House holds, are counter to the need for evaluators to question program planning and policy, and to support the active and direct inclusion of diverse inputs (House, 1991).

Stakeholder approaches, inviting various levels of involvement of participants and decisions makers, led to the transition of evaluation thinking from scientific inquiry to more illuminative use for the benefit of the program. (Papineau & Kiely, 1996) In a discussion with prominent evaluators, Patton offers an extended description of this transition:

BEST COPY AVAILABLE

> "The more scientific mode is aimed at more generalizable kinds of knowledge. The other is more situational, more situationally specific to people and to places. The scientific mode is looking for generalizable knowledge. Any specific situation is simply a place to generate information that's really relevant through generalization to the larger world. And part of the tension, then, between the researcher and the practitioner, at whatever level, whether we're talking policy or classroom, is that practitioners tend to be less interested in serving the purpose of generalization than in getting their own answers. So the researcher who's driven by the desire for generalization tends to be likely to be somewhat less responsive to practitioners situational needs, because they recognize that these needs are very situational, and won't yield as much generalizable information" (Patton, in Alkin, et al., 1990, p. 117-118).

Gold (1983) concurs with Patton that, "evaluations designed and run exclusively it the interest of "proper" research increases the probability that the results will serve mainly the interests of the research community" (p 71).

As evaluators attempted to respond to the dual call for stakeholder involvement and increased utilization, discourse centered on the nuances of these concepts: How do we recognize and define utilization? Who are the stakeholders we should involve? What should the nature of their involvement be? To what extent should they be involved?

Describing and Recognizing Utilization

Various conceptions of utilization have been proffered by evaluation theorists and practitioners. Leviton and Hughes (1981) classified utilization as either instrumental or conceptual. Instrumental use is directly related to specific decision points or judgements, while conceptual use describes the more formative, developmental aspects of evaluation use, or enlightenment (Weiss, 1977), or demystification (Berk & Rossi ,1977). Conceptual use implies that people are affected in how they think about an issue. As learning takes place both with individuals and groups, new information is incorporated with old, creating opportunities for development and refinement. (Forss, et al., 1994, p.

576). Patton defines utilization as "intended use by intended users." (Patton, in Alkin, et al., 1990, p. 192), and indicates that this requires up front negotiation with intended users as to what an evaluation can realistically achieve. He offers that utilization may involve overcoming staff fears, making sure that we are asking relevant questions, and being responsive to the situational needs of the stakeholders. Alkin (1990) characterizes utilization as the "purposeful, planned consequences that result from applying evaluation information to a problem, question, or concern at hand" (p. 19).

Numerous studies have been completed that discuss the contextual factors that contribute both to evaluation use, and non-use. Cousins and Leithwood (1986) identify policy setting factors that influence evaluation utilization. They include: information needs, decision characteristics, political climate, competing information, personal factors, and the commitment and receptiveness of the organization to evaluation. Further, Alkin, et al. (1979) points to human and context factors, such as attitudes and professional experience, financial constraints, and the relationship of the organization within the community-at-large, as important predictors of evaluation use. Preskill (1998) identifies the major problems as a mismatch between the nature of evaluation information and organizational needs. "The problem is not that there isn't known data with which to answer an organizations questions, but that the quality, timeliness and content of existing data do not meet the learning and performance information needs of organization members. Nor is sufficient time typically devoted to assigning meaning to the data that are available" (p. 5). Cox (1977) summarizes numerous factors resulting in non-use, including mismatch between the roles and styles of clients and evaluators, especially the conflicting tension between research-based models for evaluation and the developmental needs of educational practitioners working in evolving programs. He suggests that evaluation reports have been quick to point to negative outcomes without substantive suggestions for program revision and adds that lack of timely reporting, focus on irrelevant questions or issues, competing political or other issues of primary importance, and lack of a usable report have all contributed to the misuse or under-utilization of evaluation findings (p. 500).

While Preskill admits that prior research has illuminated various types of use as well as factors that contribute to use, little has been done to focus on the organizational context and culture in which use will take place (Preskill, 1994). Cox (1977) reminds evaluators of the characteristics of managers: the job pace is heavy and unrelenting, work activity is characterized by brevity, variety, and fragmentation, managers prefer action and verbal communication. Managers prefer frequent information updates even if the information is incomplete or in error. He suggests that improving evaluation utilization requires evaluator sensitivity to the organizational and political realities and communication skills of program managers.

As a result of the complexity of the cultural landscape of a diverse initiative, including varied perspectives and needs, evaluation planning and utilization may be difficult to implement and document, and may be significantly delayed as it percolates through social and political contexts. As Weiss (1981) cautions, utilization may not appear as "discrete provisions, nor in a linear sequence" (p. 18). She adds that rarely can an evaluation be comprehensive or convincing enough to supply the "correct" answer. Further, she offers that many decisions are made which do not follow rational decision processes and even when there truly are clear decisions to be made and identified decision makers, people don't always know what kind of information they need to know (Weiss, 1992, pp. 171-174). Cox (1977) adds that evaluation can focus on irrelevant questions or issues, often ignoring the important political or other context-related issues that drive decision-making.

Defining Stakeholders – Exploring the Stakeholder Approach

Various conceptions of stakeholders and stakeholder involvement evolved across the 1980's and 1990's. Weiss recognizes stakeholders as "the members of groups that are palpably affected by the program and who therefore will conceivably be affected by evaluative conclusions about the program or the members of groups that make decisions about the future of the program," and identifies four major classifications of stakeholders:

"policymakers, program managers, practitioners, and clients or citizens." (Weiss, 1983b, pp. 84-85) Gold (1983), with a similar focus on impact, identifies stakeholders as "individuals with a vested interest in the outcome of evaluations." (p. 64).

Broader conceptions of stakeholders encompass all program constituents, including funders, planners, participants and anyone with a vested interest in the program or activity as participants throughout the evaluation process, even though they may not have direct decision-making input. Likewise, the level of involvement of stakeholders may vary from consultation with a key decision maker (Cooley & Bickel, 1985) to the self-determined and conducted evaluation suggested in empowerment evaluation (Fetterman 1996, 1997). Weiss proposes that participative approaches to evaluation fall on a continuum based on the level of control the evaluator retains, with most stakeholder approaches using participants in a consultative fashion while the evaluator still maintains control, moving to more collaborative approaches with shared control, to empowerment evaluation where the evaluator is a consultant to a stakeholder controlled and designed evaluative process (Weiss, 1998, pp. 99-100).

Emerging Practice: Conceptions of Stakeholder Approaches

Despite the definitional difficulties with the identification and levels of involvement of stakeholders and the consequential changes in evaluation and evaluator roles, the participatory experience enacted through stakeholder involvement has been documented to create "opportunities for exchanges, discussion and synergy" among stakeholder cohorts (Papineau & Kiely, 1996, p. 87) and can result in evaluation that "can be both responsible and responsive to many interests." (Gold, 1983, p. 71) "Through this process [stakeholder involvement], an evaluation approach that is potentially more responsible, more realistic, and more valuable for its contribution to program development and knowledge acquisition can be achieved" (Gold, 1983, p. 72).

Patton (1997) claims that "all participatory, collaborative, and utilization-focused approaches emphasize a facilitative role for evaluators and include illuminative outcomes

for participants" (p. 149). The process of implementing stakeholder evaluations can involve diverse options based on the factors of stakeholder involvement and utilization discussed earlier. Gold (1983) proposes several general steps that form a foundation for most participative approaches. These include: determining initial stakeholder expectations for the program, negotiating a set of workable expectations within the reality of the program and its evaluation capacity, and modifying either stakeholder expectations or the program as required. He suggests that stakeholder-evaluator co-development and modification of expectations are characteristic of participative approaches and requires "an evaluation approach [that] is both dynamic and interactive" (p. 72).

Cousins and Earl (1995) distinguish participatory approaches by goals (increased use, theory generation, participant emancipation) and the degree of researcher-participant collaboration. Cousins, Donahue and Bloom (1995) identified three dimensions for categorizing evaluations: degree of researcher v. practitioner control, depth of participation, and breadth of stakeholder participation (limited primary users to all potential stakeholder cohorts). Fetterman, as explicated in his empowerment evaluation approach (1996, 1997) has added new dimensions to these existing categories, indicating the degree of enhanced self-determination, evaluator advocacy for stakeholder groups or causes, and the degree that training is an explicit goal or process of the evaluation.

Developmental Evaluation

Patton (1994) reflected on his changing role as an evaluator with some organizations, and commented that a "developmental" approach to evaluation best characterized his practice with some clients. In describing the role of evaluation, he indicated that these stakeholders "never expect to arrive at a steady state of programming because they're constantly tinkering as participants, conditions, learnings, and context change" (p. 313). He reports his use of "developmental evaluation" which he defines as "evaluation processes and activities that support program, project, product, personnel and/or organizational development (usually the latter)" (Patton, 1994, p. 314). Summarizing his recommendation for utilization focused evaluation, he states that it "shifts attention from

methods or the object of evaluation (e.g., the program) to the intended users of evaluative processes and information, and their intended uses" (Patton, 1994, p. 317). He cites the original intent of the formative-summative distinction (Scriven, 1967) suggested "that formative evaluation was meant only as a method to prepare for summative evaluation procedures" (Patton, 1994, p. 312). Over time, he believes the meaning of formative evaluation extended to include any evaluation whose "primary purpose is program improvement, where higher goal attainment was still the ultimate goal" (Patton, 1994, p. 313). He distinguishes developmental evaluation by its focus on programs or processes where continuing development *is* the outcome. In these situations, "clarity, specificity, and measurability are limiting" concepts to drive evaluative efforts. Patton relates that these participants "don't aspire to arrive at a model subject to summative evaluation and generalization. Rather, they aspire to continuous progress, ongoing adaptation and rapid responsiveness ... moreover, they don't conceive of development and change as necessarily improvements" (Patton, 1994, p. 313). Patton offers developmental evaluation as a relationship that describes the role of the participants, including the evaluator, rather than a formal model or approach.

Empowerment evaluation and evaluative inquiry are additional examples of participative stakeholder approaches to evaluation, though both offer a more formal set of guidelines or procedures than does Patton's discussion of developmental evaluation.

Empowerment Evaluation

Empowerment evaluation has philosophical roots in an emancipatory research tradition that grew out of liberation pedagogy, feminist inquiry, and critical theory (Patton, 1997). Fetterman (1997) describes empowerment evaluation as "the use of evaluation concepts, techniques, and findings to foster improvement and self-determination. It employs both qualitative and quantitative methodologies" (p. 4). He adds that it "has an unambiguous value orientation – it is designed to help people help themselves and improve their programs using a form of self-evaluation and reflection" (p. 5). Zimmerman (in press) comments that empowerment evaluation

> "redefines the professional's role relationship with the target population. The professional's role becomes one of collaborator and facilitator rather then expert and counselor. The professional's skills, interests, or plans are not imposed on the community; rather, professionals become a resource for a community. This role relationship suggests that what professionals do will depend on the particular place and people with whom they are working, rather than on technologies that are predetermined to be applied in all situations" (p. 5).

Recognizing that it is participant involvement that "empowers" stakeholders, not the evaluator, Fetterman (1997) views the approach as "necessarily a collaborative group activity, not an individual pursuit. It invites (if not demands) participation, examining issues of concern to the entire community in an open forum" (p. 5). He indicates that the approach is useful to "evaluate a situation by degrees rather than an absolute" (Fetterman, 1997, p. 379). Further, he makes the case that "merit and worth are not static values ... populations shift, goals shift, knowledge about program practices and their value changes, and external forces are highly unstable," indicating that empowerment evaluation seeks to internalize and institutionalize a strategy to more effectively merge evaluation practice with emerging needs (Fetterman, 1996, p. 6). Fetterman proposes major components of empowerment evaluation: training, facilitation, illumination and liberation, where each step includes dialogue (Fetterman, 1996). Empowerment evaluation builds on the suggestion by Cronbach (1980) that evaluators step outside a purely technical role to become an educator to share insights about the program, as well as the conduct of the evaluation. Patton (1997) suggests that the preeminence of training to build evaluation capacity is a distinguishing feature of Fetterman's approach.

Looking at potential outcomes of empowerment evaluation, Zimmerman (in press) proposes different outcomes at various levels of analysis, noting that "when we are studying organizations, outcomes might include organizational networks, effective resource acquisition, and policy leverage" (p. 381). Fetterman specifies general outcomes related to empowerment and capacity building, including, equalization of power, development of a community of learners, self-determination, and

institutionalization of evaluation practice (Fetterman, 1996).

Organizational Learning – Evaluative Inquiry

Focusing on issues of utilization, Preskill (1994, 1996, 1998) has proposed to infuse a process of evaluative inquiry within organizational cultures. She indicates that the "history of the organization, and its resultant culture, political environment and resources, provide an important context in which to view the factors that are found to directly influence evaluation use" (Preskill, 1994, p. 266). She acknowledges the influence of Watkins and Marsick (1993) who suggest that learning organizations depend on six actions: creating continuous learning opportunities, promoting inquiry and dialogue, encouraging collaboration and team learning, empowering people toward a collective vision, and connecting the organization to its environment (p. 11). Preskill and Torres (1998) propose a three phase process to foster organizational learning: focusing the inquiry, carrying out the inquiry, and applying the inquiry (p.1). They envision a process during each of the phases that includes "collective action of dialogue, reflection, asking questions, and identifying and clarifying individuals' values, beliefs, assumptions and knowledge" (p. 1). Organizational structure is important to support this process. Within the organizational infrastructure they identify the key components of culture, leadership, communication, and systems and structures that must be in place to support evaluative inquiry and organizational learning (Preskill & Torres, 1998). Preskill and Torres (1998) view evaluative inquiry as "something that engages all organizational members on a daily basis" not limited to more typical event-driven evaluation approaches (p. 7).

Preskill (1991) proposes a phenomenological approach to studying organizational culture, providing "thick descriptions" (Geertz, 1973) "that allow ambiguities, contradictions, and paradoxes to be explored with relative ease" (Siehl & Martin, 1988).
The approach also employs the use of iterative "clinical" interviews to uncover organizational culture (Schein, 1985). Preskill concludes "culture provides a necessary framework for making sense of the multiple realities that exist in every organization. It is the critical lens that helps evaluators see what strategies should be used in an evaluation

to increase its potential use" (Preskill, 1991, p. 13).

All three of the examples included in this discussion clearly articulate a changing role for both the evaluation process and the evaluator. These issues seem at the heart of stakeholder approaches, and in sharp contrast to the more "objectivist" (Stufflebeam, 1997) evaluation models developed and predominantly used in the field prior to the 1980's.

The Role of Evaluation in Stakeholder Approaches

Developmental needs of program stakeholders who become active participants in the evaluation process impact the role of evaluation within the organization. "The stakeholder approach marks a change in evaluation priorities. The salience of quantitative, summative assessment of the value of the program concept is reduced. ... such interests are jostled for dominance by demands from other stakeholders for current information on a barrage of practical questions (Weiss, 1983, p. 10-11).

Cronbach (1980) concluded that this transition would serve well, indicating that "accountability emphasizes looking back in order to assign praise or blame; evaluation is better used to understand events and processes for the sake of guiding future activities" (p. 383). The focus to help understand and better guide program components occurs through the substantive consultation and involvement of stakeholders, allowing evaluation to more routinely serve ongoing informational needs. Cooley and Bickel (1985) describe their Decision-Oriented Educational Research (DOER) approach to stakeholder evaluation, which recommends the regular involvement of key decision-makers and is "not research designed to clarify or defend particular theoretical notions but, rather, is a very applied research designed to inform the day-to-day guidance of educational systems" (p. 3).

Weiss (1983b) supports "illuminative" evaluations that produce "responsive, relevant, well-circulated results [that] can provide information that keeps people well informed

about a range of programmatic issues" (p. 91). She suggests that "without dictating specific decisions, they can permeate people's understanding of program potentials and limits" (p. 91). Mathison (1994) describes the relationship between evaluation and the developmental aspects of the program, indicating that "inserting an evaluation catalyst" into an organization leads to internalization of inquiry to support process understanding (p. 300). Preskill (1998) recommends a deep and inclusive institutionalization of evaluation suggesting that "evaluation ought to serve as an ongoing source of information to help organizational stakeholders examine and resolve issues and concerns deemed important within the organization" (p. 1). She envisions an ongoing process where "the purpose of evaluation shifts from making formative and summative decisions with evaluation being an event, to evaluation being a process that facilitates ongoing learning in organizations" (Preskill & Torres, 1996, p. 30). Torres (1994) summarizes the intent of this role of evaluation to address overall organizational issues rather than specific programs within organizations.

Weiss (1983) reminds us that "neither the political environment nor the organizational milieu is stable. Program decision-making is beset by unexpected occurrences from inside and outside the organization" (p. 87). Framed as an organizational development role, evaluation becomes an "integral, ongoing process that contributes to individual, team and organizational learning" (Preskill & Torres, 1996, p. 1). Of course, this changing role for evaluation must be supported by an organizational infrastructure and culture that is characterized "a willingness to learn and change" and that those values must be shared by the evaluator, as well (Mathison, 1994, p. 301).

Changing roles for evaluation require a different set of evaluation techniques and tools to more adequately portray and represent organizational culture and context. Weiss (1983) concludes that stakeholders are "a more reliable source of information if the evaluation is a qualitative, illuminative investigation of program operation" because most evaluations using qualitative techniques can more accurately shift direction and method as learning is revealed and new avenues become available for inquiry" (p. 88). Preskill and Torres

(1998) concur that "as continuous change becomes the normal state in organizations, evaluators will need to broaden their purpose and corresponding set of tools if they wish to have any kind of impact on organizations' success" (p. 1).

The Role of the Evaluator in Stakeholder Approaches

Patton (1997) claims "all participatory, collaborative, and utilization focused styles of evaluation change the role of the evaluator from the traditional lone judge of merit or worth to a facilitator of judgments by others involved in the process, sometimes in addition to the evaluator's judgment and sometimes without independent judgment by the evaluator." (p. 149). While the evaluator's role is "central to successful implementation of the stakeholder approach," (Gold, 1983, p. 71) many evaluators have been formally trained in evaluation techniques that lend themselves to "objectivist" evaluation designs. Patton builds on data collected through the American Evaluation Association indicating that most evaluators at this time (1988) identified themselves historically and traditionally with a primary discipline of study, and only secondarily as program evaluators. He also reports on research done by Shadish and Epstein (1988) who reported that evaluators whose primary professional identity is evaluation were more likely to resonate with stakeholder approaches, whereas those who identified primarily with an academic discipline were more likely to approach evaluation emphasizing research outcomes and summative findings (Patton, 1990). Gold (1983) concurs by summarizing that "the research procedures in which most evaluators have been trained encourage methodical, cautious, and independent behavior. Although this behavior is important, it can become counterproductive in developing useful evaluations if rigidly applied without modification or adaptability" (p. 71). Weiss (1983) emphasizes the collaborative role evaluators must play, stressing the expanded purview of their role indicating "they are asked not only to be technical experts who do competent research, they are required to become political managers who orchestrate the involvement of diverse interest groups" (p. 10). Cohen (1983) further clarifies that "this additional burden will require a keen awareness of the political context of the environment, and increased skills for the evaluator" (p. 74).

This fundamental change in the evaluator's role can present both challenge and danger. Murray (1983) summarizes the dilemma: "Done right, the stakeholder approach unavoidably thrusts the evaluator headlong into ethical and professional tugging and pulling that make the ordinary rough and tumble of the applied research environment seem quiet by comparison" (p. 60). Gold (1983) cautions "if evaluators identify too closely with the research role with which they are familiar and comfortable, parts of the stakeholder process can appear frightening" (p. 71).

What does the need for contextual sensitivity and additional skills mean for an evaluator involved in participative approaches to evaluation? As part of a collaborative team of program stakeholders, Patton (1994) describes the evaluator as

> "part of a team whose members collaborate to conceptually design and test new approaches and a long-term, ongoing process of continuous improvement, adaptation, and intentional change. The evaluator's primary function in the team is to elucidate team discussions with evaluative data and logic, and to facilitate databased decision-making in the developmental process" (p. 317).

He more specifically describes the scope of his involvement by saying "what I bring to the design team is evaluation logic, knowledge about effective programming based on evaluation wisdom, and some methods expertise to help set up monitoring and feedback systems." He further clarifies both his and other members' roles in decision-making by indicating that "all team members render evaluation judgments together and decide how to apply the implications of results for the next stage of development" (pp. 313-314).

Preskill and Torres (1996) characterize evaluation within organizational development efforts as "evaluative inquiry" that "asks evaluators to the collaborators, facilitators, interpreters, mediators, coaches and educators of learning and change processes. It asks evaluators to develop longer term relationships with organization members so that they too can become knowledgeable and skilled in evaluation theory and practice" (p. 30). Preskill (1994) relating her evaluation efforts in organizations offers specific

recommendations:

> "for evaluators engaged in facilitating organizational learning, this means redefining how we negotiate and enact our roles. For external consultants, it means spending more time not only designing and conducting evaluations, but staying with the organization to plan and implement the changes made as a result of the evaluation. For internal evaluators, it means developing stronger cross-organizational relationships and closer ties with upper management. For all evaluators it means understanding the organization's business and strategic plans and being able to access the channels of communication throughout the organization. Philosophically, it represents a shift in thinking about the purpose of evaluation and the role of objectivity and values --- from the value-free objective scientist to the "neutral-advocate" of programs, policies, and procedures. Practically, it means developing or refining another set of skills that enable us to mediate conflict, guide others in dialogue, negotiate across boundaries, understand and manage team dynamics, and work with constantly changing organizations and resulting power structures. In this sense the evaluator's job becomes a blending of the traditional evaluator and organization development consultant" (p. 296).

Brown (1995) points to the setting for the evaluation as an important determinant of evaluator role, notably, "in a social science context that acknowledges multiple perspectives and realities, it is easier to discuss the advantages and disadvantages of the evaluator as co-learner rather than expert, conveyor of information rather than deliverer of truth ... educator rather than judge" (p. 204).

Response to These Approaches From the Field

Participative stakeholder approaches represent a major shift in evaluation practice. Just as the prior models and approaches they were meant to augment carry with them limitations and issues of concern, so do stakeholder approaches. Cohen (1983) voices a major concern in stakeholder evaluation, namely, that "a single evaluation contract would

become a vehicle for managing the expression of far more values than it had been the past" (p. 74). This expanded role for evaluation and evaluators carries a burden that is well-beyond expectations of prior models, and may be unattainable or at least compromised in actual practice. The issues of stakeholder identification and level of involvement are further compromised when considering whether one evaluation can serve the variety of needs presented by stakeholder involvement (Weiss, 1983).

Weiss (1983) specifies many of the assumptions operating within participative models, including: stakeholder groups can be identified in advance, stakeholders want to be involved in the evaluation, they want specific information to inform process, evaluators will respond to stakeholders, stakeholder involvement will lead to a sense of ownership both in the evaluative process and findings, and stakeholders actually have decisions to make that evaluation can speak to. She indicates that often these assumptions are unfulfilled, and perhaps not even examined closely prior to engaging in evaluative activity. Murray (1983) reminds us that "the intense, continual personal interactions that [a stakeholder approach] requires with all the parties to an evaluation" requires "much more frequent, detailed, and affect laden [contacts] than in the usual evaluation" (p. 60). This may require skills and resources not readily available to the evaluator. He characterizes these interactions as "both its strength and its danger" (p. 56). Cohen (1983) also indicates the importance of the extent to which stakeholders are organized and articulate, another factor that may fall beyond the control or influence of the evaluator.

Stufflebeam (1997) clusters empowerment evaluation with other forms of what he calls "relativist evaluation" (p. 325) including discrepancy evaluation, responsive evaluation, naturalistic evaluation and goal-free evaluation. He characterizes the distinguishing feature of relativistic evaluations as the validation of criteria of worth and merit primarily on the endorsement of some interested party (p. 11). He suggests that Fetterman (empowerment evaluation) and other proponents of "relativistic" evaluation have "fallen prey to a key logical flaw that Scriven has identified ... confusing the potential roles of

an evaluation with its essential nonvariant goal of determining something's value, or subordinating that goal by a focus on the processes that lead to the determination of value" (p. 326). He admonishes readers to beware that this approach could be used as a "cloak of legitimacy" to redirect awareness from the key issues of evaluation (Stufflebeam, 1997, p. 324). Weiss also questions what types of evaluative inquiry are most compatible with stakeholder involvement and concludes that evidence would suggest the use of qualitative, formative techniques, thus limiting the scope of numerous useful evaluation methods (Weiss, 1983). Patton (1994) admits that his use of "developmental evaluation" requires an understanding that "crossing that line, however, does reduce independence of judgment. The costs and benefits of such a role change must be openly acknowledged and carefully assessed" (p 316).

Stufflebeam (1997) extends his critique by contrasting "relativist" approaches with "objectivist" evaluation that is "based on the theory that moral good is objective and independent of personal or merely human feelings" (p. 326). According to him, objectivist evaluations are:

> "firmly grounded in ethical principles, strictly control bias or prejudice in seeking determinations of worth or merit, invoke and justify appropriate and (where they exist) established standards of merit and worth, obtain and validate findings from multiple sources, set forth and justify conclusions about the evaluand's merit and/or worth, report findings honestly and fairly to all right-to-know audiences, and subject the evaluation process and findings to independent assessments against the standards of the evaluation field" (p. 326).

Stufflebeam believes that the processes of evaluation training, reflection and self-determination embodied in empowerment evaluation (as well as many participative, collaborative and stakeholder approaches), while worthy activities, are not evaluation, and suggests that presuming they comprise evaluation, will do a grave disservice to the field (Stufflebeam, 1997, pp. 326-327). These criticisms squarely land in the realm of paradigmatic perspectives. While Stufflebeam relies on a world view of objective truth,

many stakeholder approaches are deeply linked with interpretive or critical theory perspectives. This "debate" is not only instructive and interesting as academic discourse within the field of evaluation, but it is also of critical importance in practice. Matching evaluator and evaluand perspectives, as well as the subsequent purpose and conduct of the evaluation, is crucial to avoid misunderstandings and unrealistic expectations. Stakeholder approaches may not be acceptable to external agencies responsible for funding or ultimately seeking accountability audits. Reflecting this awareness of multiple audiences for evaluation, Murray (1983) indicates that a stakeholder approach "unavoidably pushes the evaluation toward technical compromises that lead to diminished long term gains in knowledge," that can be generalized to the wider population (p. 60).

Gold (1983) offers that stakeholder input is useful in "specifying the kinds of evaluative information required; and the most useful form for presentation of the information" (p. 64). This involvement can facilitate a better exchange between stakeholders and evaluators to determine the forms of evaluation, and the respective traditions of inquiry, that can best serve specific settings and needs. Gold conceptualizes sustained input by stakeholders through periodic feedback by the evaluator. He sees this dialogue as a two way street, providing formative program information to stakeholders, and an ongoing check and balance system for the evaluator (Gold, 1983, p. 64).

Fetterman counters Stufflebeam's critique by asserting that empowerment evaluation is conducted by a community of learners who determine the scope, focus, methods and use for the evaluation and offers that through community representation, bias is explicitly apparent, and resolved or dealt with appropriately (Fetterman, 1997, p. 183). He also reminds us that assuming an objectivist stance does not necessarily insure the control of bias, "often overlooking the integral prejudice of background, training, and perspective the evaluator brings to the evaluation" (Fetterman, 1997, p. 185). In determining the best forms of evaluation for use within specific contexts "the focus should be on the problem or issue; methods and methodologies should follow, not precede" (Fetterman, 1997,

p.188). Fetterman mirrors the sentiments of Weiss when he characterizes evaluation, "like any other dimension of life, [it] is political, social, cultural, and economic. It rarely produces a single truth or conclusion" (Fetterman, 1997, p. 188).

Evaluation as Discursive Practice

While the three participative approaches discussed here, and many other variations of stakeholder evaluation in practice, differ in key areas, they share common elements, as well. Obviously, they all share some level of involvement by stakeholders, though the purpose of the involvement, the identification of stakeholders, and the levels of involvement may differ. These factors also determine the role of the evaluation and the evaluator within the process. The nature of the stakeholder participation involves some form of dialogue between stakeholders and evaluator, with more participative, collaborative, developmental, or empowerment approaches involving a considerable degree of dialogue and collaborative decision-making. The dialogue seeks to create new knowledge as well as a deeper understanding of existing data. It is this deeper form of dialogue that represents collective reflection and action that I characterize as *discursive practice*. In addition, many participative approaches are educative, both in the experience of the evaluation process, but also, in building internal evaluation capacity.

Evaluation based on discursive practice recognizes that "organizations learn through joint discussion and interpretation of events, and through gradual changes in the assumptions, symbols, and values of participants. In this approach, trials and errors, or actions and outcomes, are important means of learning" (Daft & Huber, 1987, p. 10). This requires a perspective that views the organization as a "living entity that can disassemble, recombine, quickly respond to internal and external stimuli, and build organization members' self direction and self organizing capacities" (Preskill & Torres, 1996, p. 3). Senge (1990) proposes that a learning organization is where people "continually expand their capacity to create the results they truly desire, where new and expansive patterns of thinking are nurtured, where collective aspiration is set free, and where people are continually learning how to learn together." (p. 3).

BEST COPY AVAILABLE

Discursive practice includes reflection, the generation and sharing of new data and knowledge, and iterative cycles of deliberation:

> "as people engage in dialogue in a public context around a shared problem ... data are generated. As these data are organized and presented to them as ... a problematic situation upon which they can reflect and act, dialogue is further experienced. Interactors are able to stand back from their experience, include others' perspectives of the situation in their own, and recognize themes, patterns, and contradictions in their shared context. More data are generated and analyzed. This iterated cycle of data generated and analysis through dialogue leads to the development of a common language and shared understanding of the situation, and a transformation of the system" (Hazen, 1986, August, p. 4).

Torres (1994) links this activity directly to individual and organizational learning through shared dialogue. She identifies this key process as "the articulation of issues and concerns in dialogue with others that facilitates individuals' internalization of improvement efforts. This interaction constitutes the key linkage between individual and organizational learning" (p. 334).

Building internal evaluation capacity, participative approaches can assist organizations in developing a culture of inquiry "where information collection, development and utilization becomes a matter of fact, ongoing activity for all those working in the organization" (Bhola, 1995, p. 11). Torres, Preskill & Piontek (1996) "see the most realistic role for evaluative information as one that contributes in an evolutionary way to both understanding and decisions" (p. 48). They also make a useful distinction between discussion and dialogue:

> "while the terms discussion and dialogue are often used interchangeably, we believe there are differences worth noting. For example, the purpose of discussion is to tell, sell, or persuade. It is an attempt to find agreement, defend one's assumptions, or convince someone of an idea. Dialogue, however, seeks to inquire, to share meanings, to understand the whole, and to uncover one's

assumptions. Discussion is about individuals and preserving the status quo; dialogue is about communities and learning for change" (p. 21).

Senge, Roberts, Ross, Smith & Kleiner (1994) provide another useful description of dialogue as

> "a sustained collective inquiry into everyday experience and what we take for granted. The goal of dialogue is to open to ground by establishing a "container" or "field" for inquiry, a setting where people can become more aware of the context around their experience, and of the processes of thought and feeling that created that experience" (p. 353).

It is the integral use of stakeholder dialogue, through iterative deliberation, that characterizes discursive practice; a combination of active reflection and the creation of shared knowledge and understanding. Through this form of participation, evaluation can support organizational development and fuller evaluation utilization, while concurrently, building an institutionalized evaluation capacity.

References

Alkin, M. C., Burry, J., & Ruskus, J. (1984) The administrator's role in evaluation use. ERIC Document Reproduction Service No. ED 251 515, November.

Alkin, M. C. (1990). Debates on evaluation. Newbury Park, CA: Sage.

Alkin, M. C., Daillak, R. & White, P. (1979). Using evaluation: Does evaluation make a difference? Beverly Hills, CA: Sage.

Allegheny Policy Council. (1994). A formula for the future: A regional approach to math and science education. Pittsburgh, PA: The Allegheny Policy Council.

Barabba, V. P., & Zaltman, G. (1991) Hearing the voice of the market: Competitive advantage through creative use of market information. Boston, MA: Harvard Business School Press.

Berk, R. A. & Rossi. P. H. (1977). Doing good or worse: Evaluation research politically reexamined. Evaluation studies review annual, 2, 77-89.

Bhola, H. S. (1995). Informed decisions within a culture of information: Updating a model of information development and evaluation. Adult education for development, 44, 75-85.

Bloom, B. S., Engelhart, M.D., Furst, E. J., Hill, W. H. & Krathwohl, D. R. (1956). Taxonomy of educational objectives: Handbook I: Cognitive domain. New York: David McKay.

Brown, J. (1995). Dialogue: Capacities and stories. In, S. Chawla & J. Renesch (Eds.), Learning organizations: Developing cultures for tomorrow's workplace (pp. 153-164). Portland, OR: Productivity Press.

Bunt, N. R. (1997). The early evolution of SW PA's regional math/science collaborative from the leadership perspective. Unpublished doctoral dissertation, University of Pittsburgh, Pittsburgh.

Campbell, D. T. (1984). Can we be scientific in applied social science? In R. F. Connor, D. G. Altman, & C. Jackson (Eds.), Evaluation studies review annual, 9. Beverly Hills, CA: Sage.

Cohen, D. K. (1983). Evaluation and reform. In A. S. Bryk (Ed.), Stakeholder Based Evaluation. San Francisco:Jossey-Bass.

Coleman, J. S. & Campbell, D. T. (1966). Equality of educational opportunity. Washington, D.C.: U.S. Government Printing Office.

Comer, J. "The Education of Inner City Children." Grants Magazine, 1980, 1, 20-26.

Cooley, W. W. & Bickel, W. E. (1985). Decision-oriented educational research. Boston: Kluwer-Nijhoff.

Cornbach, L. J., & Suppes, P. (1969). Research for tomorrow's schools: Disciplined inquiry for education. New York: Macmillan.

Cousins, J. B. & Earl, L. M. (1995). The case for participatory evaluation: Theory, research, practice. In J. B. Cousins & L. M. Earl (Eds.), Participatory evaluation in education: Studies in evaluation use and organizational learning. Washington, D. C.: Falmer Press.

Cousins, J. B. & Leithwood, K. A. (1986). Current empirical research on evaluation utilization. Review of Educational Research, 58. pp. 331-364.

Cousins, J. B., Donohue, J. J., & Bloom, G. (1996). Understanding collaborative evaluation: Results from a survey of North American evaluators. Unpublished manuscript, University of Ottawa.

Cronbach, L. J., et al. (1980). Toward reform of program evaluation . San Francisco: Jossey-Bass.

Cronbach, L. J. (1963). Course improvement through evaluation. Teachers college record, 64, 672-683.

Daft, R. L. & Huber, G. P. (1987). How organizations learn: A communication framework. Research in the sociology of organizations, 5, 1-36.

Edmonds, R. (1979). "Effective Schools for the Urban Poor." Educational Leadership. 37, 115-27.

Eisner, E. W. (1976). Educational connoisseurship and criticism: Their form and function in educational evaluation. Journal of Aesthetic Education, 10, 135-150.

Fetterman, D. M. (1995). A response to Stufflebeam's review of empowerment evaluation. Evaluation practice, 16, (2) 179-199.

Fetterman, D. M. (1996). Conclusion: Reflections on emergent themes and next steps. In D. M. Fetteman, S. J. Kaftarian, & A. Wandersman(Eds.), Empowerment evaluation: Knowledge and tools for self-assessment & accountability. (pp. 379-385). Thousands Oaks, CA: Sage .

Fetterman, D. M. (1997). A response to Patton and Scriven's review of empowerment evaluation. Evaluation practice, 18, (3) 253-266.

Fetterman, D. M.. Kaftarian, S. J., & Wandersman, A. (Eds.). (1996). Empowerment evaluation: Knowledge and tools for self-assessment & accountability. Thousands Oaks, CA: Sage .

Fischer, C. et al. (1980). "Teaching Behaviors, Academic Learning Time and Student Achievement: An Overview" In Time to Learn. Edited by C. Denham and A. Lieberman. Washington, D.C.: U.S. Office of Education.

Forss, K., Cracknell, B., & Samset, K. (1994). Can evaluation help an organization to learn? Evaluation review, 18, (5) 574-591.

Gold, N. (1983). Stakeholders and program evaluation: Characterizations and reflections. In A. S. Bryk (Ed.), Stakeholder-based evaluation. (pp. 63-72). San Francisco: Jossey-Bass.

Guba, E. G. & Lincoln, Y. S. (1986) Do inquiry paradigms imply methodologies? In D. M. Fetterman (ed.), The silent scientific revolution. Beverly Hills, CA: Sage.

Guba, E. G. & Lincoln, Y. S. (1989). Fourth generation evaluation. Newbury Park, CA: Sage.

Guba, E. G. & Lincoln, Y. S. (1981) Effective evaluation: Improving the usefulness of evaluation results through responsive and naturalistic approaches. San Francisco: Jossey-Bass.

House, E. R. (1980). Evaluating with validity. Beverly Hills, CA: Sage.

House, E. R. (1983). Assumptions underlying evaluation models. In G. F. Madaus, M. Scriven & D. L. Stufflebeam (Eds.), Evaluation models: Viewpoints on educational and human services evaluation. Boston: Kluwer-Nijhoff.

House, E. R. (1991). Evaluation and social justice: Where are we? In M. McLaughlin & D. Phillips (Eds.), Evaluation and education at quarter century, National Society for the Study of Education Yearbook. (pp. 233-246). Chicago: University of Chicago Press.

House, E. R. (1993). Professional evaluation. Beverly Hills, CA: Sage.

Joint Committee on Standards for Educational Evaluation. (1994). The Program Evaluation Standards. 2nd Ed. Thousand Oaks, CA: Sage.

Krathwohl, d. R., Bloom, B. S., & Masia, B. B. (1964). Taxonomy of educational objectives: Handbook II: Affective domain. New York: David McKay.

Leviton, L. C. & Hughes, E. F. X. (1981). Research on the utilization of evaluations: A review and synthesis. Evaluation Review, 5. pp. 525-548.

BEST COPY AVAILABLE

Lincoln, Y. S. (1991). The arts and sciences of program evaluation. Evaluation Practice, 12, (1) 1-7.

Madaus, G. F., Scriven, M. & Stufflebeam, D. L. (1983). Evaluation models: Viewpoints on educational and human services evaluation. Boston: Kluwer-Nijhoff.

Mathison, S. (1994). Rethinking the evaluation role: Partnerships between organizations and evaluators. Evaluation and Program Planning, 17(3) pp. 299-304.

Murray, C. A. (1983). Stakeholder as deck chairs. In A. S. Bryk (Ed.), Stakeholder Based Evaluation. San Francisco: Jossey-Bass.

Papineau, D. & Kiely, M. C. (1996). Participatory evaluation in a community organization: Fostering stakeholder empowerment and utilization. Evaluation and Program Planning, 19(1) pp. 79-93.

Patton, M. Q. (1997b). Utilization-focused evaluation: The new century text (3rd ed.). Thousand Oaks, CA: Sage .

Patton, M. Q. (1986). Utilization-focused evaluation. Beverly Hills, CA: Sage.

Patton, M. Q. (1988a). How primary is your identity as an evaluator? Evaluation Practice, 9, (2), 87-92.

Patton, M. Q. (1988b). The evaluator's responsibility for utilization. Evaluation Practice, 9, (2) 5-24.

Patton, M. Q. (1994). Developmental evaluation. Evaluation practice, 15, (3) 311-319.

Patton, M. Q. (1997a). Toward distinguishing empowerment evaluation and placing it in a larger context. Evaluation Practice, 18, (2) 147-163.

Patton, M. Q. (1998, April 20) Re: Reflecting on our past. American Evaluation Association Discussion List [Online]. Available E-mail: EVALTALK@UA1VM.UA.EDU [1998, April 20].

Preskill, H. (1991). The cultural lens: Bringing utilization into focus. In Larson, C. L. & Preskill, H. (Eds.), Organizations in transition: Opportunities and challenges for evaluation,. no. 49. (pp. 5-15). San Francisco: Jossey-Bass Inc., Publishers.

Preskill, H. (1994). Evaluation's role in enhancing organizational learning: A model for practice. Evaluation and Program Planning,17, (3), 291-297.

Preskill, H. & Torres, R. T. (1996, November). From evaluation to evaluative inquiry for organizational learning. Paper presented at the American Evaluation Association annual conference, Atlanta, GA.

Preskill, H. & Torres, R. T. (1998, March). Building capacity for organizational learning through evaluative inquiry. Paper presented at the meeting of the Academy of Management's mid-year conference on Organizational Learning, Washington, D.C.

Qualitative evaluation and research methods. (1990). Newbury Park, CA: Sage.

Regional Math/Science Collaborative. (1996-1997). Journal, 2, 1-79.

Rice, J. M. (1915). The people's government: Efficient, bossless, graftless. Philadelphia: John C. Winston.

Rich, R. E. (1977). Use of knowledge by federal bureaucrats. In C. H. Weiss (Ed.), Using social research in public policy making. (pp. 193-211). Lexington, MA: Heath.

Schein, E. H. (1985). Organizational culture and leadership. San Francisco: Jossey-Bass.

Scriven, M. (1967) The methodology of evaluation. In R. E. Stake (Ed.), Curriculum evaluation. American Educational Research Association Monograph Series on Evaluation, No. 1. Chicago: Rand-McNally.

Scriven, M. (1984). Evaluation ideologies. In R. F. Connor, D. G. Altman, & C. Jackson (Eds.), Evaluation studies review annual. (Vol. 9). Beverly Hills, CA: Sage.

Scriven, M. (1998, April 20) Re: Reflecting on our past. American Evaluation Association Discussion List [Online]. Available E-mail: EVALTALK@UA1VM.UA.EDU [1998, April 20].

Senge, P. (1990). The fifth discipline. New York: Doubleday.

Senge, P. M., Roberts, C., Ross, R. B., Smith, B. J., & Kleiner, A. (1994). The fifth discipline fieldbook. New York: Doubleday.

Shadish, W. R. (1998, April 20) Re: Reflecting on our past. American Evaluation Association Discussion List [Online]. Available E-mail: EVALTALK@UA1VM.UA.EDU [1998, April 20].

Shadish, W. R., Jr., & Epstein, R. (1987). Patterns of program evaluation practice among members of the Evaluation Research Society and Evaluation Network. Evaluation Review, 11, 555-590.

Shoemaker, J. (1986). "Developing Effectiveness in the District, School and Classroom." Equity and Choice.

Should or can evaluation be goal-free?. (1972). Evaluation comment: The journal of educational evaluation (Center for the study of evaluation), 3, (4), 7-9.

Siehl, C. & Martin, J. (1988). Measuring organizational culture: Mixing qualitative and quantitative methods. In M. Owens, M. Moore, & R. C. Snyder (Eds.), Inside organizations. Newbury Park, CA: Sage.

Smith, J. K. & Heshusius, L. (1986). Closing down the conversation: The end of quantitive-qualitative debate. Educational Researcher, 15, (1), 4-12.

Stake, R. E. (1969). Generalizability of program evaluation: The need for limits. Education Product Report. 2(5), pp. 39-40.

Stake, R. E. (1998, April 20) Re: Reflecting on our past. American Evaluation Association Discussion List [Online]. Available E-mail: EVALTALK@UA1VM.UA.EDU [1998, April 20].

Stufflebeam, D. L. (1994). Empowerment evaluation, objectivist evaluation, and evaluation standards: Where the future of evaluation should not go and where it needs to go. Evaluation practice, 15, (3) 321-338.

Stufflebeam, D. L. (1969). Evaluation as enlightenment for decision making. In W. H. Beatty & A. B. Walcott (Eds.), Improving educational assessment and an inventory of measures of affective behavior. Washington, D.C.: Association for Supervision and Curriculum Development.

Stufflebeam, D. L. (1971). The relevance of The CIPP Evaluation Model for Educational Accountability. Journal of research and development in education, 5(1), 19-25.

Stufflebeam, D. L. (1983). The CIPP model for program evaluation. In G. F. Madaus, M. Scriven, & D. L. Stufflebeam (Eds.), Evaluation models: Viewpoints on educational and human services evaluation. Boston: Kluwer-Nijoff.

Talmage, H. (1982). Evaluation of programs. In H. E. Mitzel (Ed.), Encylcopedia of educational research (5th ed.). New York: The Free Press.

Tananis, C. A. (1998, April 20) Reflecting on our past. American Evaluation Association Discussion List [Online]. Available E-mail: EVALTALK@UA1VM.UA.EDU [1998, April 20].

Torres, R. T. (1994). Linking individual and organizational learning: The internalization

and externalization of evaluation. Evaluation and program planning, 17(3). pp. 327-337.

Torres, R. T., Preskill, H. S., & Piontek, M. E. (1996). Evaluation strategies for communicating and reporting: Enhancing learning in organizations. Thousands Oaks, CA: Sage.

Travers, R. M. W. (1983). How research has changed American schools. Kalamazoo, MI: Mythos Press.

Tyler R. W. (1942). General statement on evaluation. Journal of educational research, 35, pp. 492-501.

Watkins, K. E. & Marsick, V. J. (1993). Sculpting the learning organizations. San Francisco: Jossey-Bass.

Weiss, C. H. (1981). Measuring the use of evaluation. In J. A. Ciarlo (Ed.), Utilizing evaluation: Concept and measurement techniques. (pp. 17-33). Beverly Hills, CA: Sage.

Weiss, C. H. (1983a). The stakeholder approach to evaluation: Origins and promise. In A. S. Bryk (Ed.), Stakeholder-based evaluation. (pp. 3-14). San Francisco: Jossey-Bass.

Weiss, C. H. (1983b). Toward the future of stakeholder approaches in evaluation. In A. S. Bryk (Ed.), Stakeholder-based evaluation. (pp. 83-96). San Francisco: Jossey-Bass.

Weiss, C. H. (1988). Evaluation for decisions: Is anybody there? Does anybody care? Evaluation practice, 9, (1) 5-19.

Weiss, C. H. (1998). Evaluation: Methods for studying programs and policies (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

Worthen, B. R. & Sanders, J. R. (1987). Educational Evaluation: Alternative Approaches and Practical Guidelines. White Plains, N.Y.: Longman.

Zimmerman, M. A. *(in press)*. Empowerment theory: Psychological, organizational, and community levels of analysis. In J. Rappaport & E. Seldman (eds.), Handbook of community psychology. New York: Plenum.

BEST COPY AVAILABLE